

# Quantum Approximate Markov Chains are Thermal

Kohtaro Kato

*Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo, Japan*

Fernando G.S.L. Brandão

*Institute for Quantum Information and Matter  
California Institute of Technology, Pasadena, CA 91125, USA*

We prove an upper bound on the conditional mutual information of Gibbs states of one-dimensional short-range quantum Hamiltonians at finite temperature. We show the mutual information between two regions  $A$  and  $C$  conditioned on a middle region  $B$  decays exponentially with the square root of the length of  $B$ . Conversely, we also prove that any one-dimensional quantum state with small conditional mutual information in all tripartite splits of the line can be well-approximated by a Gibbs state of a local quantum Hamiltonian. These two results constitute a variant for one-dimensional quantum systems of the Hammersley-Clifford theorem (which characterizes Markov networks, i.e. probability distributions which have vanishing conditional mutual information, as Gibbs states of classical Hamiltonians). The result can be seen as a strengthening – for one-dimensional systems – of the mutual information area law for thermal states. It directly implies a method to efficiently prepare any one-dimensional Gibbs state at finite temperature by a constant-depth quantum circuit.

## I. INTRODUCTION

A sequence of random variables  $X_1, \dots, X_n$  forms a Markov chain if  $X_{i+1}$  is uncorrelated from  $X_1, \dots, X_{i-1}$  conditioned on the value of  $X_i$ . Markov chains are a central concept in probability theory, statistics and beyond. In this paper we consider two natural generalizations of the concept of a Markov chain.

In the first we only require approximate independence from previous random variables, i.e.  $X_{i+1}$  should only be *almost* independent from  $X_1, \dots, X_{i-1}$  conditioned on  $X_i$ . One way to make this notion quantitative is to use the conditional mutual information, defined for every three random variables  $X, Y, Z$  drawn from the distribution  $p(X, Y, Z)$  as

$$I(X : Z|Y)_p := H(XY)_p + H(YZ)_p - H(XYZ)_p - H(Y)_p, \quad (1)$$

with  $H(X)_p := -\sum_{x_i \in X} p_X(x_i) \log p_X(x_i)$  the Shannon entropy of the marginal distribution on  $X$ . The conditional mutual information can also be written as

$$I(X : Z|Y)_p = \mathbb{E}_{Y \sim p(Y)} I(X : Z)_{p_y}, \quad (2)$$

with  $p_y(X, Z)$  the conditional distribution of  $X$  and  $Z$  given  $Y = y$ . Thus if  $I(X : Z|Y)_p$  is small,  $X$  and  $Z$  are almost uncorrelated conditioned on  $Y$ . We say  $X_1, \dots, X_n$  is a  $\varepsilon$ -approximate Markov chain if

$$I(X_1, \dots, X_{i-1} : X_{i+1}, \dots, X_n | X_i) \leq \varepsilon \quad \forall i \in [n]. \quad (3)$$

In the second, instead of considering random variables distributed according to a probability distribution  $p(X_1, \dots, X_n)$ , we consider a  $n$ -partite *quantum* state given by a density matrix  $\rho_{A_1, \dots, A_n} \in \mathcal{D}(\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n})$ <sup>1</sup>. The quantum conditional mutual information of a tripartite

<sup>1</sup>  $\mathcal{D}(\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n})$  is the set of density matrices over the vector space  $\mathcal{H}_{A_1} \otimes \dots \otimes \mathcal{H}_{A_n}$ .

state  $\rho_{ABC}$  is defined as

$$I(A : C|B)_\rho := S(AB)_\rho + S(BC)_\rho - S(ABC)_\rho - S(B)_\rho, \quad (4)$$

with  $S(X)_\rho := -\text{Tr}(\rho_X \log \rho_X)$  the von Neumann entropy of subsystem  $X$ . Quantum states satisfying  $I(A : C|B)_\rho = 0$  are analogues of Markov chains of three variables in the sense that there is a choice of a projective measurement in the  $B$  system for which the post-measured state on  $A$  and  $C$  is uncorrelated (i.e. product) for every possible outcome [1]. A multipartite quantum state  $\rho_{A_1, \dots, A_n}$  is a quantum Markov chain if

$$I(A_1, \dots, A_{i-1} : A_{i+1}, \dots, A_n | A_i)_\rho = 0 \quad \forall i \in [n]. \quad (5)$$

The object of study in this paper is a combination of both generalizations. We are interested in *quantum approximate* Markov chains. Such concept is already non-trivial for tripartite quantum states  $\rho_{ABC}$ . We can say  $\rho_{ABC}$  forms a quantum  $\varepsilon$ -approximate Markov chain if

$$I(A : C|B)_\rho \leq \varepsilon. \quad (6)$$

However there is no quantum analogue of Eq. (2) [2] and therefore it is unclear if the definition above is meaningful. A recent result in quantum information theory reveals its meaning [3]. It shows that

$$I(A : C|B)_\rho \geq \min_{\Delta: B \rightarrow BC} -2 \log F(\rho_{ABC}, \Delta^{B \rightarrow BC}(\rho_{AB})), \quad (7)$$

where the minimum is over all quantum channels  $\Delta$  mapping  $\mathcal{D}(\mathcal{H}_B)$  to  $\mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$ , and  $F(\rho, \sigma) := \text{Tr}((\sigma^{1/2} \rho \sigma^{1/2})^{1/2})$  is the fidelity. Thus if the conditional mutual information is small,  $A$  is only correlated to  $C$  through  $B$  up to a small error, in the sense that  $C$  can be approximately recovered given the information contained in  $B$  only. We say  $\rho_{A_1, \dots, A_n}$  is a quantum  $\varepsilon$ -approximate Markov chain if

$$I(A_1, \dots, A_{i-1} : A_{i+1}, \dots, A_n | A_i)_\rho \leq \varepsilon \quad \forall i \in [n]. \quad (8)$$

### A. The Hammersley-Clifford Theorem

In this paper we will be interested in finding a structural characterization of quantum approximate Markov chains. Our motivation is a powerful result in statistics called Hammersley-Clifford Theorem [4]. It states that Markov chains (and more generally Markov networks<sup>2</sup>), in which all elements of the distribution are non-zero, are equivalent to the set of Gibbs (thermal) states of 2-local (nearest-neighbor) Hamiltonians:

$$p(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \exp \left( - \sum_i h_i(x_i, x_{i+1}) \right), \quad (9)$$

for functions  $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ , with

$$Z := \sum_{x_1, \dots, x_n} \exp \left( - \sum_i h_i(x_i, x_{i+1}) \right) \quad (10)$$

<sup>2</sup> A Markov network is a generalization of a Markov chain given by random variables  $X_{V_1}, \dots, X_{V_k}$  defined on the vertices  $V_1, \dots, V_k$  of a graph, such that  $X_{V_i}$  is uncorrelated from all other random variables conditioned on the random variables  $\{X_{V_j}\}_{j \sim i}$  associated to neighboring vertices.

the partition function<sup>3</sup>.

In Refs. [5, 6], the Hammersley-Clifford theorem was generalized to quantum Markov chains (and Markov networks): A full rank quantum state  $\rho_{A_1 \dots A_n}$  is a quantum Markov chain if, and only if, it can be written as

$$\rho = \frac{\exp(-\sum_i h_{i,i+1})}{\text{Tr}(\exp(-\sum_i h_{i,i+1}))}, \quad (11)$$

where each  $h_{i,i+1}$  only acts on subsystems  $A_i, A_{i+1}$ , such that  $[h_{i,i+1}, h_{j,j+1}] = 0$  for all  $i, j$ . Therefore we have a characterization of full rank quantum Markov chains as Gibbs states of 1D commuting local quantum Hamiltonians<sup>4</sup>.

The characterization above only involves a very special set of quantum Hamiltonians. A natural question is whether there is a setting for which we get the set of all 1D quantum Gibbs states. The main result of this paper answers the question in the affirmative: we prove that quantum approximate Markov chains are equivalent to Gibbs states of one-dimensional local quantum Hamiltonians.

**Notation:** In the following, we consider a 1D quantum spin system  $\Lambda$  with  $n$  spins. We assume that each local subsystem  $\mathcal{H}_i$  corresponding to spin  $i$  has the finite dimension  $d < \infty$ . We denote the operator norm of an operator  $V$  by  $\|V\|$ , and the trace norm of  $V$  by  $\|V\|_1$ . We say that the support of  $V$  is  $X$  if an operator  $V$  on  $\Lambda$  can be written as

$$V = V_X \otimes \text{id}_{X^c}, \quad (12)$$

i.e., the tensor product of some non-trivial operator  $V_X$  on the spins in a region  $X$  and the identity operator acting on (the spins in) the complement of  $X$  (which we denoted by  $X^c$ ). We will denote the support of an operator  $V$  by  $\text{supp}(V)$ .

Consider a disjoint tripartition  $ABC$  of  $\Lambda$ , where  $A$  and  $C$  are indirectly connected through  $B$ . We denote by  $d(A, C)$  the minimum distance between  $A$  and  $C$ . If  $B$  is connected,  $d(A, C) = |B|$ . We consider short-range Hamiltonians in  $\Lambda$  given by  $H_{ABC} = \sum_i h_i$ , where each  $h_i$  is bounded and  $\text{supp}(h_i)$  acts on  $r < \infty$  neighboring spins around the spin  $i$ . We also consider a restricted Hamiltonian  $H_X$  on a region  $X \subset \Lambda = ABC$ , defined as

$$H_X = \sum_{\text{supp}(h_i) \subset X} h_i, \quad (13)$$

i.e., the sum of interactions acting on spins sitting *inside* of  $X$ . The Gibbs state  $\rho^{H_X}$  of the Hamiltonian  $H_X$  is defined as

$$\rho^{H_X} = \frac{e^{-\beta H_X}}{Z_X}, \quad (14)$$

where  $Z_X = \text{Tr}[e^{-\beta H_X}]$  is the partition function. The reduced state of the Gibbs state on a subregion  $Y \subset X$  is denoted by  $\rho_Y^{H_X}$ .

<sup>3</sup> For Markov networks, in turn, the Hamiltonian is a sum of local functions of variables on all cliques of the graph.

<sup>4</sup> In Ref. [6] a more general result was shown for quantum Markov networks. In contrast to the classical case, positive (i.e. full rank) quantum Markov networks are only equivalent to Gibbs states of commuting local Hamiltonians with terms on the cliques of the graph if the graph is triangle free.

## II. RESULTS

In this section we present the main results of this paper. Let  $l_0, C, c > 0$  be universal constants (which can be taken to be less than 100). Our first result is the following theorem (see Section III for the proof):

**Theorem 1.** *Let  $H = \sum_i h_i$  be a short-range one-dimensional Hamiltonian with  $\|h_i\| \leq 1$ . For an inverse temperature  $\beta > 0$ , let  $\rho_{H,\beta} := e^{-\beta H} / \text{Tr}(e^{-\beta H})$ . Then for every tripartite split of the lattice  $ABC$ , there exist a quantum channel  $\Lambda_{B \rightarrow BC} : \mathcal{H}_B \rightarrow \mathcal{H}_B \otimes \mathcal{H}_C$  such that*

$$\|\rho_{H,\beta} - \Lambda_{B \rightarrow BC}(\text{tr}_C(\rho_{H,\beta}))\|_1 \leq e^{-q(\beta)\sqrt{d(A,C)}} \quad (15)$$

for any  $d(A, C) \geq l_0$ , with  $q(\beta) = Ce^{-c\beta}$ .

The theorem above states that if we choose the region  $B$  sufficiently large, the Gibbs state can be approximately recovered from the partial trace over  $C$  by performing a recovery map on  $B$ . In turn, the statement implies the corresponding conditional mutual information decays similarly:

**Corollary 2.** *Under the setting of Theorem 1,*

$$I(A : C|B)_{\rho_{H,\beta}} \leq \left( d(A, C) + 3q(\beta)^{-2}d(A, C)^{-1/2} \right) e^{-q(\beta)\sqrt{d(A,C)}}. \quad (16)$$

Our second main result is a converse to Theorem 1. For two quantum states  $\rho, \sigma$ , define their relative entropy as

$$S(\rho||\sigma) := \text{Tr}(\rho(\log \rho - \log \sigma)). \quad (17)$$

if  $\text{supp}(\rho) \subset \text{supp}(\sigma)$ , and  $S(\rho||\sigma) := +\infty$  otherwise (Here “ $\text{supp}(\rho)$ ” is the subspace spanned by eigenvectors of  $\rho$  with nonzero eigenvalues).

In Section IV we prove:

**Theorem 3.** *Let  $\rho_{A_1, \dots, A_n}$  be an quantum  $\varepsilon$ -approximate Markov chain. Then there exists a local Hamiltonian  $H = \sum_i h_{A_i, A_{i+1}}$ , with  $h_{A_i, A_{i+1}}$  only acting on  $A_i A_{i+1}$ , such that*

$$S\left(\rho \left\| \frac{e^{-H}}{\text{Tr} e^{-H}}\right.\right) \leq \varepsilon n. \quad (18)$$

We note that the result is only non-trivial when  $\varepsilon < 1/n$ . From the relation

$$S(\rho||\sigma) \geq -2 \log F(\rho, \sigma), \quad (19)$$

we find that the state  $\rho$  is also close in fidelity to a Gibbs state of a local model.

Theorem 1 shows that any 1D Gibbs state of a short-range Hamiltonian is a quantum approximate Markov chain. Conversely, Theorem 3 shows that any 1D state which is a quantum approximate Markov chain can be well-approximated by 1D Gibbs state of some short-range Hamiltonian. Therefore the combination of the two results gives as a variant of the Hammersley-Clifford theorem for quantum approximate Markov chains. Below we discuss two implications of the results:

### A. Saturation Rate of Area Law for 1D Gibbs States

In Ref. [7], it is shown that the Gibbs state of a short-range Hamiltonian obeys an area law in terms of the mutual information. For instance, for any Gibbs state  $\rho$  of a short-range Hamiltonian on a lattice, it holds that [7]

$$I(A : A^c)_\rho \leq 2\beta J |\partial A|, \quad (20)$$

where  $J$  is a constant only depending on the locality of  $H$  and the norm of the local interactions. When the system is one-dimensional the upper bound is a constant.

Consider a one-dimensional spin chain. Let  $A$  be some connected region, and  $B_l := b_1 \dots b_l$  be a region composed of sites  $b_i$  surrounding  $A$ . Due to the monotonicity of mutual information under the partial trace,

$$I(A : B_l)_\rho \leq 2\beta J |\partial A| \quad (21)$$

holds for any sufficiently large  $l$ . Corollary 2 goes one step further and quantifies the *rate* at which  $I(A : B_l)_\rho$  saturates when  $l$  grows. Indeed:

$$I(A : B_{l+1}) - I(A : B_l) = I(A : b_{l+1} | B_l) \leq C l e^{-c\sqrt{l}}, \quad (22)$$

for constants  $C, c > 0$ .

Therefore, the mutual information of a 1D Gibbs state saturates quickly with an increasing region  $B$ .

### B. A Short Depth Representation of 1D Gibbs states

Theorem 1 also has the following consequence:

**Corollary 4.** *The Gibbs state of any local Hamiltonian in one dimension can be well-approximated by a depth-two (mixed) circuit with each gate acting on  $O(\log^2(n))$  qubits.*

*In more detail, there is a quantum operation of the form*

$$\Delta = \left( \bigotimes_i \Delta_{2,i} \right) \left( \bigotimes_i \Delta_{1,i} \right), \quad (23)$$

*with each quantum channel  $\Delta_{k,i}$  acting on  $O(e^{O(\beta)} \log^2(n/\varepsilon))$  sites, with  $\Delta_{k,i}$  and  $\Delta_{k,j}$  acting on non-overlapping sites for  $i \neq j$ , such that*

$$\left\| \Delta(\tau) - \frac{e^{-\beta H}}{\text{Tr}(e^{-\beta H})} \right\|_1 \leq \varepsilon, \quad (24)$$

*with  $\tau$  the maximally mixed state.*

An earlier result [8] proved that 1D Gibbs states of local Hamiltonians at finite constant temperature can be approximated by a matrix product operator of polynomial bond dimension, which implies they can be constructed efficiently on a quantum computer. However this result does not give that the state can be constructed by a short depth quantum circuit, as Corollary 4 shows.

### C. Higher Dimensions: A Conjecture

The quantum Hammersley-Clifford theorem holds for more general class of Markov networks such as regular  $D$ -dimensional lattices. In this case, a partition  $ABC$  of the system is chosen so that  $A$  and  $C$  are separated by  $B$  (Figure 1). We say  $B$  shields  $A$  from  $C$ . Due to this observation, we expect that Gibbs states of short-ranged Hamiltonians obey fast decay of the conditional mutual information in  $D$ -dimensional systems as well. We have the following conjecture:

**Conjecture 1.** *Let  $\rho$  be a Gibbs state of a short-ranged Hamiltonian defined on a  $D$ -dimensional spin lattice. Then, there exist constants  $C, c > 0$  such that for every three regions  $A, B, C$  with  $B$  shielding  $A$  from  $C$ ,*

$$I(A : C|B)_\rho \leq C e^{-cd(A,C)} \quad (25)$$

We will discuss a partial result towards the conjecture in Sec. III E. Note that when  $D = 1$ , this conjecture gives an improved bound. It turns out that the area law for mutual information implies a weak version of the conjecture. Consider a Gibbs state in the infinite volume limit (as a KMS state). Let  $A$  be a region of the lattice and  $B_l$  be a ring around  $A$  of width  $l$ . Then we know that for every  $l$  [7]:

$$I(A : B_l) \leq c\beta|\partial A|. \quad (26)$$

By the data processing inequality,  $I(A : B_l)$  is monotonically increasing with  $l$ . Therefore for every  $\varepsilon > 0$  there is an integer  $l$  s.t.

$$I(A : B_{l+1}) - I(A : B_l) \leq \varepsilon, \quad (27)$$

which can be written as

$$I(A : C|B_l) \leq \varepsilon, \quad (28)$$

with  $C := B_{l+1} \setminus B_l$ .

We can also ask whether Theorem 3 can be extended to higher dimensions, i.e. is any state on a  $D$ -dim lattice with small  $I(A : C|B)$  for  $B$  shielding  $A$  from  $C$  close to thermal? Although we do not know any counter-example, we also could not find any partial result in this direction.

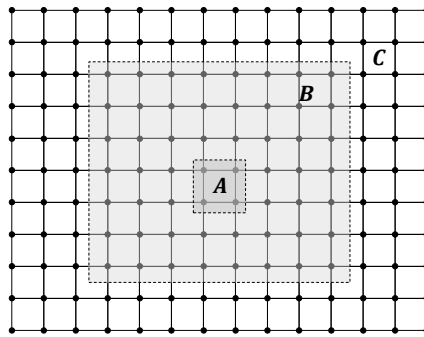


FIG. 1: An example of a 2D lattice with a partition  $ABC$ . We expect that the conditional mutual information  $I(A : C|B)_\rho$  for any Gibbs state decays fast with respect to  $d(A, C)$ .

### III. APPROXIMATE MARKOV PROPERTY OF 1D QUANTUM GIBBS STATES

In this section, we provide a proof of Theorem 1. A key point of the proof is that if a short-range Hamiltonian changes locally, the corresponding Gibbs state also changes locally. To obtain operators representing changes of the Gibbs state, we employ quantum belief propagation equations which have been studied in [9, 10]. We discuss another approach based on the results of Ref. [11] in Appendix A.

#### A. Perturbative analysis of Gibbs states

Let us consider an one parameter family of a Hamiltonian  $H$  under a perturbation by  $V$

$$H(s) = H + sV, \quad (29)$$

where  $0 \leq s \leq 1$ . The effect on the Gibbs state due to a small change of  $s$  can be computed through a differential equation given by [9, 10]

$$\frac{d}{ds} e^{-\beta H(s)} = -\frac{\beta}{2} \left\{ e^{-\beta H(s)}, \Phi_\beta^{H(s)}(V) \right\}. \quad (30)$$

Here, the operator  $\Phi_\beta^{H(s)}(V)$  is defined by

$$\Phi_\beta^{H(s)}(V)_{ij} := V_{ij} \tilde{f}_\beta(E_i(s) - E_j(s)) \quad (31)$$

in energy eigenbasis of the Hamiltonian  $H(s) = \sum_i E_i(s) |i\rangle\langle i|$  (each  $|i\rangle$  depends on  $s$  as well as  $E_i(s)$ ), where  $\tilde{f}_\beta(\omega) = \frac{\tanh(\beta\omega/2)}{\beta\omega/2}$ . By using the Fourier transform  $f_\beta(t) = \frac{1}{2\pi} \int d\omega \tilde{f}_\beta(\omega) e^{i\omega t}$ ,  $\Phi_\beta^{H(s)}(V)$  is written as the integral of filtered time-evolutions of  $V$ , that is,

$$\Phi_\beta^{H(s)}(V) = \int_{-\infty}^{\infty} dt f_\beta(t) e^{-iH(s)t} V e^{iH(s)t}. \quad (32)$$

By computing the solution of Eq. (30), we obtain that

$$e^{-\beta H(1)} = O e^{-\beta H(0)} O^\dagger, \quad (33)$$

where the operator  $O$  is defined as

$$O = \mathcal{T} \text{Exp} \left[ -\frac{\beta}{2} \int_0^1 ds' \Phi_\beta^{H(s')}(V) \right] \quad (34)$$

$$= \sum_{n=0}^{\infty} \left( -\frac{\beta}{2} \right)^n \int_0^1 ds'_1 \int_0^{s'_1} ds'_2 \cdots \int_0^{s'_{n-1}} ds'_n \Phi_\beta^{H(s'_n)}(V) \cdots \Phi_\beta^{H(s'_1)}(V), \quad (35)$$

with  $\mathcal{T}$  the time-ordering operation. Since we have  $dt f_\beta(t) = \frac{dt}{\beta} f_1(\frac{t}{\beta})$ , it holds that

$$\left\| \Phi_\beta^{H(s)}(V) \right\| = \left\| \int_{-\infty}^{\infty} dt' f_1(t') e^{-i\beta H(s)t'} V e^{i\beta H(s)t'} \right\| \quad (36)$$

$$\leq \|V\| \left| \int_{-\infty}^{\infty} dt' f_1(t') \right| = \|V\|. \quad (37)$$

The integral in the last equality can be calculated through the series expansion of  $\tanh(x)/x$ , which is written as

$$\frac{\tanh(x)}{x} = \sum_{k=0}^{\infty} \frac{2}{x^2 + (k + \frac{1}{2})^2 \pi^2}. \quad (38)$$

The upper bound of  $\|\Phi_{\beta}^{H(s)}(V)\|$  implies the upper bound of the operator norm of  $O$  that is given by

$$\|O\| \leq e^{\frac{\beta}{2}\|V\|}. \quad (39)$$

Since the Hamiltonian defined on a many-body system is short-ranged, we can restrict the time-evolution of a local operator by using the Lieb-Robinson bound [12]. Suppose that  $H$  is a Hamiltonian obeying the Lieb-Robinson bound, and  $O_A$  and  $O_B$  are observables supported on local regions  $A$  and  $B$ , respectively. One convenient representation of the Lieb-Robinson bound for these operators is formulated as follows.

$$\|[O_A, e^{-iHt} O_B e^{iHt}]\| \leq c \|O_A\| \|O_B\| \min(|A|, |B|) e^{c'(vt - d(A, B))}, \quad (40)$$

where  $c, v \geq 0, c' > 0$  are constants. Assume that  $H(0)$  is a short-range Hamiltonian and  $\text{supp}(V)$  is some connected local region. Then  $H(s)$  obeys the Lieb-Robinson bound for all  $0 \leq s \leq 1$ . Since  $f_{\beta}(t)$  decays fast in  $|t|$ , the Lieb-Robinson bound implies that the effective support of  $\Phi_{\beta}^{H(s)}(V)$  can be restricted to a local region  $\mathcal{V}_l$ , which contains all sites within some distance  $l$  from  $\text{supp}(V)$ . More precisely, there exist positive constants  $c'$  and  $v$ , which is determined by  $H(s)$ , such that [10]

$$\|\Phi_{\beta}^{H(s)}(V) - \text{Tr}_{\mathcal{V}_l^c} [\Phi_{\beta}^{H(s)}(V)] \otimes \mathbb{I}_{\mathcal{V}_l^c}\| \leq c' \|V\| e^{-\frac{c'l}{1+c'v\beta/\pi}}. \quad (41)$$

We also define the integral of the restricted operator in Eq. (41) as

$$O_{\mathcal{V}_l} := \mathcal{T} \text{Exp} \left[ -\frac{\beta}{2} \int_0^1 ds' \text{Tr}_{\mathcal{V}_l^c} [\Phi_{\beta}^{H(s')}(V)] \otimes \mathbb{I}_{\mathcal{V}_l^c} \right]. \quad (42)$$

This operator is also localized on  $\mathcal{V}_l$  and approximates  $O$  with good accuracy. Let us choose  $c'$  and  $v$  so that Eq. (41) holds for all  $0 \leq s \leq 1$ . Then, it holds that

$$\|O - O_{\mathcal{V}_l}\| \leq \frac{c'\beta\|V\|}{2} e^{\frac{(1+c')\beta\|V\|}{2}} e^{-\frac{c'l}{1+c'v\beta/\pi}}. \quad (43)$$

To prove this, let us first consider some parametrized operators  $Q(s)$  and  $\tilde{Q}(s)$  satisfying  $\|Q(s)\|, \|\tilde{Q}(s)\| \leq C$ . Then, it holds that

$$Q(s_n)Q(s_{n-1}) \cdots Q(s_1) = \tilde{Q}(s_n)\tilde{Q}(s_{n-1}) \cdots \tilde{Q}(s_1) \quad (44)$$

$$+ \sum_{j=1}^n Q(s_n) \cdots Q(s_{j+1}) \Delta_j \tilde{Q}(s_{j-1}) \cdots \tilde{Q}(s_1) \quad (45)$$

where  $\Delta_j = Q(s_j) - \tilde{Q}(s_j)$  and  $\|\Delta_j\| \leq \Delta$ . Therefore, we obtain that

$$\|Q(s_n)Q(s_{n-1}) \cdots Q(s_1) - \tilde{Q}(s_n)\tilde{Q}(s_{n-1}) \cdots \tilde{Q}(s_1)\| \leq nC^{n-1}\Delta. \quad (46)$$



Let us choose  $Q(s) = \Phi_\beta^{H(s)}(V)$  and  $\tilde{Q}(s) = \text{Tr}_{\mathcal{V}_l^c} [\Phi_\beta^{H(s)}(V)] \otimes \mathbb{I}_{\mathcal{V}_l^c}$ . From Eq. (37) and Eq. (41), their norms are bounded as

$$\|Q(s)\|, \|\tilde{Q}(s)\| \leq \|Q(s)\| + \|\tilde{Q}(s) - Q(s)\| \quad (47)$$

$$\leq \left(1 + c' e^{-\frac{c'l}{1+c'v\beta/\pi}}\right) \|V\| \quad (48)$$

$$\leq (1 + c') \|V\|. \quad (49)$$

Therefore, Eq. (46) holds for  $C = (1 + c') \|V\|$  and  $\Delta = c' e^{-\frac{c'l}{1+c'v\beta/\pi}} \|V\|$ . By inserting Eq. (46) to the definitions of  $O$ , we obtain Eq. (43).

## B. The proof of Theorem 1

For the convenience of the reader, we restate Theorem 1 below.

**Theorem 1** *Let  $H = \sum_i h_i$  be a short-range one-dimensional Hamiltonian with  $\|h_i\| \leq 1$ . For an inverse temperature  $\beta > 0$ , let  $\rho_{H,\beta} := e^{-\beta H} / \text{Tr}(e^{-\beta H})$ . Then for every tripartite split of the lattice  $ABC$ , there exist a constant  $l_0 > 0$  and a quantum channel  $\Lambda_{B \rightarrow BC} : \mathcal{H}_B \rightarrow \mathcal{H}_B \otimes \mathcal{H}_C$  such that*

$$\|\rho_{H,\beta} - \Lambda_{B \rightarrow BC}(\text{tr}_C(\rho_{H,\beta}))\|_1 \leq e^{-q(\beta)\sqrt{d(A,C)}}$$

for any  $d(A, C) \geq l_0$ , with  $q(\beta) = \Omega(e^{-\Theta(\beta)})$ .

The proof of Theorem 1 consists of three steps. In the first step, we show that there exists a CP-map which recovers the Gibbs state from the reduced state with exponentially good accuracy. In the second, we normalize the CP-map to make it trace non-increasing i.e. we show the existence of a quantum operation which succeed to recover the Gibbs state with some probability. We also show that the success probability is a constant of system size. Finally, we construct a CPTP-map from the probabilistic operation by employing a repeat-until-success strategy in the third step.

Let us begin with the following lemma.

**Lemma 5.** *For any 1D Gibbs state  $\rho^{H_{ABC}}$  of a short-range Hamiltonian on a system with a partition  $ABC$  such that  $l := d(A, C)/2 > r$ , there exists a CP map  $\kappa_{B \rightarrow BC}$ , non-negative constants  $c'$  and  $v$  such that*

$$\|\rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 \leq C_1(\beta) e^{-q_1(\beta)l}, \quad (50)$$

where  $C_1(\beta)$  is a non-negative constant and  $q_1(\beta) = \frac{c'}{1+c'v\beta/\pi}$ , as defined in Eq. (41).

*Proof.* Without loss of generality, we consider in the case of  $r = 2$ , i.e., nearest-neighbor interactions. Assume that  $d(A, C) = |B|$  and  $|B|$  is even. We split region  $B$  into left hand side  $B^L$  and right hand side  $B^R$  so that each part has size  $l$  and touches only either  $A$  or  $C$ . We define  $H_{B^M} = h_{i,i+1}$  as the term in the Hamiltonian which acts on both  $B^L$  and  $B^R$ . We denote  $\|H_{B^M}\| = J$ .

**Remark 6.** *For general short-range Hamiltonians,  $d(A, C)$  should be sufficiently large so that no interaction term in  $H$  acts on both  $A$  and  $C$ . Then, we define  $H_{B^M}$  as the sum of all terms acting on both  $B^L$  and  $B^R$ , and after that, the same argument is applicable. When  $B$  consists of a fixed number of connected regions, each connected component which is neighboring both  $A$  and  $C$  is divided into two parts as in the same way. Then,  $H_{B^M}$  is the sum of all terms acting on both such divided regions.*

Consider a parametrized Hamiltonian  $H_{ABC}(s) = H_{ABL} + H_{BRC} + sH_{BM}$ . Clearly,  $H_{ABC}(1) = H_{ABC}$  and  $H_{ABC}(0) = H_{ABL} + H_{BRC}$ . Then, we introduce an operator  $O_{ABC}$  defined as

$$O_{ABC} := \mathcal{T}\text{Exp} \left[ -\frac{\beta}{2} \int_0^1 ds' \Phi_\beta^{H_{ABC}(s')} (H_{BM}) \right]. \quad (51)$$

Eq. (33) implies that

$$e^{-\beta H_{ABC}} = O_{ABC} e^{-\beta(H_{ABL} + H_{BRC})} O_{ABC}^\dagger. \quad (52)$$

Therefore,  $O_{ABC}$  adds the interaction corresponding to  $H_{BM}$  to  $\rho^{H_{ABL} + H_{BRC}}$ . We also introduce the inverse operator  $\tilde{O}_{ABC}$ , that is given by

$$\tilde{O}_{ABC} := \tilde{\mathcal{T}}\text{Exp} \left[ \frac{\beta}{2} \int_0^1 ds' \Phi_\beta^{H_{ABC}(s')} (H_{BM}) \right], \quad (53)$$

where  $\tilde{\mathcal{T}}$  denotes the inverse time-ordering operator.  $O_{ABC}$  and  $\tilde{O}_{ABC}$  satisfies the following relation (see e.g., [13])

$$O_{ABC} \tilde{O}_{ABC} = \tilde{O}_{ABC} O_{ABC} = \mathbb{1}_{ABC}. \quad (54)$$

As Eq. (39) the operator norms of  $O_{ABC}$  and  $\tilde{O}_{ABC}$  can be bounded as

$$\|O_{ABC}\|, \|\tilde{O}_{ABC}\| \leq e^{\frac{\beta}{2} \|H_{BM}\|} = e^{\frac{\beta}{2} J}. \quad (55)$$

Importantly, the upper bound is independent of size of  $A$ ,  $B$  and  $C$ . From the definitions of  $O_{ABC}$  and its inverse, it is not difficult to check that the following identity holds.

$$\rho^{H_{ABC}} = O_{ABC} \left[ \text{Tr}_{BRC} \left( \tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) \otimes \rho_{BRC}^{H_{BRC}} \right] O_{ABC}^\dagger. \quad (56)$$

From Eq. (43), we know that there exist operators  $O_B$  and  $\tilde{O}_B$  whose supports are restricted on  $B$  and satisfy

$$\|O_{ABC} - O_B\|, \|\tilde{O}_{ABC} - \tilde{O}_B\| \leq \frac{c' \beta J}{2} e^{\frac{(1+c')\beta J}{2}} e^{-q_1(\beta)l}, \quad (57)$$

where  $q_1(\beta) = \frac{c'}{1+c'v\beta/\pi}$  for non-negative constants  $c'$  and  $v$  which are chosen as in Eq. (43). Let us denote

$$K(\beta) = \frac{c' \beta J}{2} e^{\frac{(1+c')\beta J}{2}}. \quad (58)$$

Then, the operator norm of the local operators  $O_B$  and  $\tilde{O}_B$  can be bounded by

$$\|O_B\|, \|\tilde{O}_B\| \leq \|O_{ABC}\| + \|O_{ABC} - O_B\| \leq e^{\frac{\beta}{2} J} + K(\beta), \quad (59)$$

which is independent of the size of  $B$ . Let  $\tilde{O}_{B|B}$  be the non-trivial part of  $\tilde{O}_B$  acting on  $B$ , i.e.,

$$\tilde{O}_B = \tilde{O}_{B|B} \otimes \mathbb{1}_{AC}. \quad (60)$$

We then define a CP-map  $\kappa_{B \rightarrow BC}$  by replacing  $O_{ABC}(\tilde{O}_{ABC})$  by local operators  $O_B(\tilde{O}_B)$  and removing partial trace over  $C$  in Eq. (56), i.e.,

$$\kappa_{B \rightarrow BC}(\sigma_B) \equiv O_B \left[ \text{Tr}_{B^R} \left( \tilde{O}_{B|B} \sigma_B \tilde{O}_{B|B}^\dagger \right) \otimes \rho_{BRC}^{H_{BRC}} \right] O_B^\dagger. \quad (61)$$

Let us denote

$$X_1 = \text{Tr}_{B^{RC}} \left( \tilde{O}_{ABC} \rho^{H_{ABC}} \tilde{O}_{ABC}^\dagger \right) \otimes \rho_{B^{RC}}^{H_{B^{RC}}}, \quad (62)$$

and

$$X_2 = \text{Tr}_{B^R} \left( \tilde{O}_{B|B} \rho_{AB}^{H_{ABC}} \tilde{O}_{B|B}^\dagger \right) \otimes \rho_{B^{RC}}^{H_{B^{RC}}}. \quad (63)$$

For any state  $\sigma_{ABC}$ , Eq. (57) implies that

$$\|X_1 - X_2\| = \left\| \text{Tr}_{B^{RC}} \left( \tilde{O}_{ABC} \sigma_{ABC} \tilde{O}_{ABC}^\dagger \right) \otimes \rho_{B^{RC}}^{H_{B^{RC}}} - \text{Tr}_{B^R} \left( \tilde{O}_{B|B} \sigma_{AB} \tilde{O}_{B|B}^\dagger \right) \otimes \rho_{B^{RC}}^{H_{B^{RC}}} \right\|_1 \quad (64)$$

$$= \left\| \text{Tr}_{B^{RC}} \left( \tilde{O}_{ABC} \sigma_{ABC} \tilde{O}_{ABC}^\dagger \right) - \text{Tr}_{B^{RC}} \left( \tilde{O}_B \sigma_{ABC} \tilde{O}_B^\dagger \right) \right\|_1 \quad (65)$$

$$\leq \left\| \tilde{O}_{ABC} \sigma_{ABC} \tilde{O}_{ABC}^\dagger - \tilde{O}_B \sigma_{ABC} \tilde{O}_B^\dagger \right\|_1. \quad (66)$$

Here we used the monotonicity of the trace-norm in the last inequality. To address the calculation, we use the following spacial case of the Hölder's inequality:

$$\|AB\|_1 \leq \|A\|_1 \|B\|. \quad (67)$$

It implies that

$$\left\| \tilde{O}_{ABC} \sigma_{ABC} \tilde{O}_{ABC}^\dagger - \tilde{O}_B \sigma_{ABC} \tilde{O}_B^\dagger \right\|_1 \quad (68)$$

$$\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \sigma_{ABC} \tilde{O}_{ABC}^\dagger \right\|_1 + \left\| \tilde{O}_B \sigma_{ABC} (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\|_1 \quad (69)$$

$$\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \right\| \left\| \sigma_{ABC} \tilde{O}_{ABC}^\dagger \right\|_1 + \left\| \tilde{O}_B \sigma_{ABC} \right\|_1 \left\| (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\| \quad (70)$$

$$\leq \left\| (\tilde{O}_{ABC} - \tilde{O}_B) \right\| \left\| \tilde{O}_{ABC}^\dagger \right\| + \left\| \tilde{O}_B \right\| \left\| (\tilde{O}_{ABC}^\dagger - \tilde{O}_B^\dagger) \right\| \quad (71)$$

$$\leq 2K(\beta) \left( e^{\frac{\beta J}{2}} + K(\beta) \right) e^{-q_1(\beta)l}. \quad (72)$$

We can bound the difference between the original Gibbs state  $\rho^{H_{ABC}}$  and  $\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})$  as follows.

$$\left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 = \left\| O_{ABC} X_1 O_{ABC}^\dagger - O_B X_2 O_B^\dagger \right\|_1 \quad (73)$$

$$\leq \left\| O_{ABC} X_1 O_{ABC}^\dagger - O_B X_1 O_B^\dagger \right\|_1 + \left\| O_B (X_1 - X_2) O_B^\dagger \right\|_1 \quad (74)$$

$$\leq \left\| (O_{ABC} - O_B) X_1 O_{ABC}^\dagger \right\|_1 + \left\| O_B X_1 (O_{ABC}^\dagger - O_B^\dagger) \right\|_1 + \|(X_1 - X_2)\|_1 \left\| O_B^\dagger \right\|^2 \quad (75)$$

$$\leq \|O_{ABC} - O_B\| \|\tilde{O}_{ABC}\|^2 \|O_{ABC}\| + \|O_{ABC}^\dagger - O_B^\dagger\| \|\tilde{O}_{ABC}\|^2 \|O_B\| + \|(X_1 - X_2)\|_1 \left\| O_B^\dagger \right\|^2 \quad (76)$$

$$\leq K(\beta) \left( e^{\frac{3\beta J}{2}} + e^{\beta J} \left( e^{\frac{\beta J}{2}} + K(\beta) \right) + 2 \left( e^{\frac{\beta J}{2}} + K(\beta) \right)^3 \right) e^{-q_1(\beta)l} \quad (77)$$

$$\leq 4K(\beta) \left( e^{\frac{\beta J}{2}} + K(\beta) \right)^3 e^{-q_1(\beta)l}. \quad (78)$$

Here we used the fact that  $\|X_1\|_1 \leq \|\rho^{H_{ABC}}\|_1 \|\tilde{O}_{ABC}\|^2 = \|\tilde{O}_{ABC}\|^2$  in the fourth line. Choosing  $C_1(\beta) = 4K(\beta) \left( e^{\frac{\beta J}{2}} + K(\beta) \right)^3$  completes the proof.  $\square$

Unfortunately, the map  $\kappa_{B \rightarrow BC}$  is in general not a trace non-increasing map, and therefore it is not physically meaningful. Thus, we normalize  $\kappa_{B \rightarrow BC}$  to make it CP and trace non-increasing map, i.e. a probabilistic quantum operation. We show that the success probability of the normalized map is given by a constant of system sizes.

**Lemma 7.** *Under the setting of lemma 5, there exists a CP and trace non-increasing map  $\tilde{\Lambda}_{B \rightarrow BC}$  such that for any  $l \geq l_0(\beta) \equiv \left\lceil \frac{\log C_1(\beta)+1}{q_1(\beta)} \right\rceil = \mathcal{O}(\beta^2)$ , it holds that*

$$\left\| \rho^{H_{ABC}} - \frac{\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]} \right\|_1 \leq C_2(\beta) e^{-q_1(\beta)l}, \quad (79)$$

where  $C_2(\beta) = \frac{2C_1(\beta)}{(1-e^{-1})}$ . Moreover,  $p = \text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]$  is a strictly positive constant which is independent of the size of subsystems  $A$ ,  $B$  and  $C$ .

*Proof.* The maximum eigenvalues  $\lambda_{\max}^{O_B}$  and  $\lambda_{\max}^{\tilde{O}_B}$  of  $O_B^\dagger O_B$  and  $\tilde{O}_B^\dagger \tilde{O}_B$  can be bounded by their operator norms from above. From Eq. (59), the upper bound of these eigenvalues is given by

$$\lambda_{\max}^{O_B}, \lambda_{\max}^{\tilde{O}_B} \leq (e^{\frac{\beta J}{2}} + K(\beta))^2. \quad (80)$$

Denote the product  $\lambda_{\max}^{O_B} \lambda_{\max}^{\tilde{O}_B}$  by  $\lambda_{\max}$ . Clearly, it has the upper bound given by

$$\lambda_{\max} \leq (e^{\frac{\beta J}{2}} + K(\beta))^4. \quad (81)$$

We define a CP and trace non-increasing map  $\tilde{\Lambda}_{B \rightarrow BC}$  as

$$\tilde{\Lambda}_{B \rightarrow BC}(\sigma_B) = \frac{1}{\lambda_{\max}} \kappa_{B \rightarrow BC}(\sigma_B). \quad (82)$$

We choose  $l_0(\beta)$  as  $l_0(\beta) = \left\lceil \frac{\log C_1(\beta)+1}{q_1(\beta)} \right\rceil$  so that  $C_1(\beta) e^{-q_1(\beta)l} \leq e^{-1} < 1$  for any  $l \geq l_0(\beta)$ . We note that  $e^{-1}$  can be replaced by any real number which is strictly smaller than 1 by choosing different  $l_0(\beta)$ . From Eq. (50), the success probability  $p$  can be bounded by

$$p := \text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})] = \left\| \tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \quad (83)$$

$$= \frac{1}{\lambda_{\max}} \left\| \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \quad (84)$$

$$\geq \frac{1}{\lambda_{\max}} \left| \left\| \rho^{H_{ABC}} \right\|_1 - \left\| \rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \right| \quad (85)$$

$$\geq \frac{1}{\lambda_{\max}} \left( 1 - C_1(\beta) e^{-q_1(\beta)l} \right) \quad (86)$$

$$\geq \frac{1}{\lambda_{\max}} \left( 1 - \frac{1}{e} \right) \quad (87)$$

$$\geq \frac{1 - e^{-1}}{(e^{\frac{\beta J}{2}} + K(\beta))^4} > 0. \quad (88)$$

The approximation error of the normalized map can be bounded as follows.

$$\left\| \rho^{H_{ABC}} - \frac{\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\text{Tr}[\tilde{\Lambda}_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})]} \right\|_1 = \left\| \rho^{H_{ABC}} - \frac{\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \right\|_1 \quad (89)$$

$$\leq \left| 1 - \frac{1}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \right| \|\rho^{H_{ABC}}\|_1 + \frac{1}{\|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1} \|\rho^{H_{ABC}} - \kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 \quad (90)$$

$$\leq \frac{C_1(\beta)e^{-q_1(\beta)l}}{1 - e^{-1}} + \frac{1}{1 - e^{-1}} C_1(\beta)e^{-q_1(\beta)l} \quad (91)$$

$$\leq \frac{2C_1(\beta)}{1 - e^{-1}} e^{-q_1(\beta)l}. \quad (92)$$

In the third line, we used the fact that

$$\left| \|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 - 1 \right| \leq C_1(\beta)e^{-q_1(\beta)l}. \quad (93)$$

which follows from

$$1 - C_1(\beta)e^{-q_1(\beta)l} \leq \|\kappa_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 \leq 1 + C_1(\beta)e^{-q_1(\beta)l}. \quad (94)$$

Thus, we conclude that lemma 7 holds by choosing  $C_2(\beta) = \frac{2C_1(\beta)}{1 - e^{-1}}$ .  $\square$

We have shown that there is a probabilistic quantum operation which approximately recovers a Gibbs state from its reduced state. We are now in a position to prove Theorem 1. Let us assume that  $d(A, C) = |B| = 3l^2 - l$  for some  $l \in \mathbb{N}$ . We divide  $B$  into  $B = B_l \bar{B}_{l-1} B_{l-1} \dots \bar{B}_1 B_1$ , so that  $B_l$  is touching with  $A$ ,  $B_1$  is touching with  $C$ , and for each  $i$ ,  $|B_i| = 2l$  and  $|\bar{B}_i| = l$  (see Fig. 2). Let us denote  $\tilde{\Lambda}_{B_i \rightarrow B_i \bar{B}_{i-1} \dots B_1 C}$ , which is introduced in lemma 7, by  $\tilde{\Lambda}_i$  and  $\text{Tr}_{\bar{B}_i B_{i+1} \dots B_1 C}$  by  $\text{Tr}_i$ . Since each  $\tilde{\Lambda}_i$  is a CP and trace non-increasing map, we introduce another CP-map  $\tilde{E}_i$  so that  $\Lambda_i = \tilde{\Lambda}_i + \tilde{E}_i$  be a CPTP-map. Then, we define CPTP-map  $\Lambda_{B \rightarrow BC}$  as

$$\Lambda_{B \rightarrow BC}(\sigma_B) = \tilde{\Lambda}_1(\sigma_B) + \left( \tilde{\Lambda}_2 + \left( \tilde{\Lambda}_3 + \left( \dots \left( \tilde{\Lambda}_l + \tilde{E}_l \right) \text{Tr}_{l-1} \tilde{E}_{l-1} \right) \dots \right) \text{Tr}_2 \tilde{E}_2 \right) \text{Tr}_1 \tilde{E}_1(\sigma_B). \quad (95)$$

Note that the map  $\left( \tilde{\Lambda}_2 + \left( \tilde{\Lambda}_3 + \left( \dots \left( \tilde{\Lambda}_l + \tilde{E}_l \right) \text{Tr}_{l-1} \tilde{E}_{l-1} \right) \dots \right) \text{Tr}_2 \tilde{E}_2 \right)$  is also a CPTP-map. Due to the assumption of translational-invariance and lemma 7, we obtain that

$$p := \text{Tr}[\tilde{\Lambda}_1(\rho_{AB}^{H_{ABC}})] = \text{Tr}[\tilde{\Lambda}_2(\rho_{AB_l \dots B_2}^{H_{ABC}})] = \dots = \text{Tr}[\tilde{\Lambda}_l(\rho_{AB_l}^{H_{ABC}})] > 0 \quad (96)$$

$$1 - p := \text{Tr}[\tilde{E}_1(\rho_{AB}^{H_{ABC}})] = \text{Tr}[\tilde{E}_2(\rho_{AB_l \dots B_2}^{H_{ABC}})] = \dots = \text{Tr}[\tilde{E}_l(\rho_{AB_l}^{H_{ABC}})] \quad (97)$$

When we input  $\rho_{AB}^{H_{ABC}}$  to  $\Lambda_{B \rightarrow BC}$ , the output of each map  $\tilde{\Lambda}_i$  corresponds to the success of the recovery process at  $i$ th step (with probability  $p$ ) and  $\tilde{E}_i$  corresponds to the failure of the recovery process (with probability  $1 - p$ ). If it fails, we trace out both the recovered system and the buffer system  $\bar{B}_i$ , and then, the effect of the failure can be almost neglected. This is guaranteed by the following lemma, which utilizes the exponential decay of correlation of 1D Gibbs states [11].

**Lemma 8.** *There exists a constant  $\xi \geq 0$  such that*

$$(1 - p) \left\| \text{Tr}_i(\rho^{H_{ABC}}) - \frac{\text{Tr}_i \tilde{E}_i(\rho_{AB_l \dots B_i}^{H_{ABC}})}{(1 - p)} \right\|_1 \leq e^{-l/\xi}. \quad (98)$$

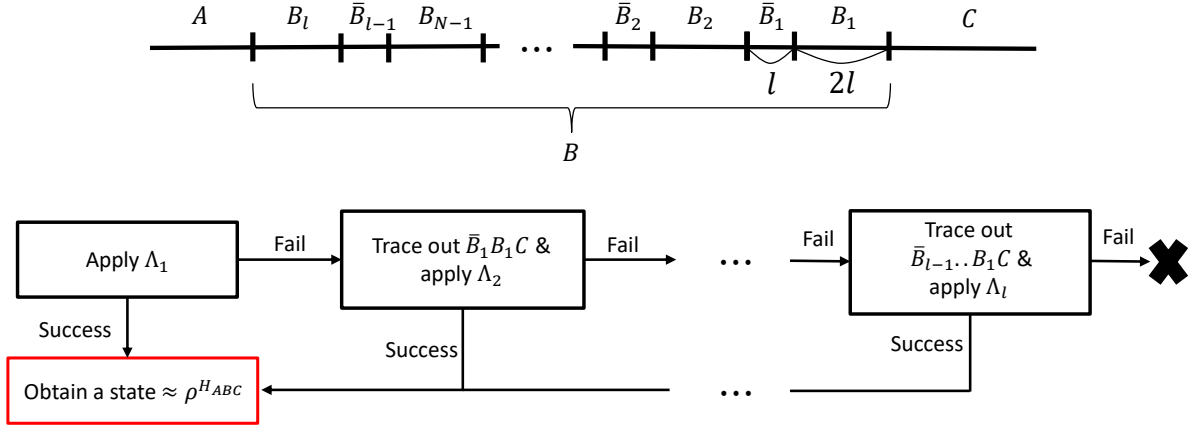


FIG. 2: A schematic picture of the repeat-until-success method. The “failure” output at the end corresponds to the CP-map  $\tilde{E}_l \circ \text{Tr}_{l-1} \tilde{E}_{l-1} \cdots \circ \text{Tr}_1 \tilde{E}_1$ .

*Proof.* Define a correlation function  $\text{Cor}(X : Y)_\rho$  of regions  $X$  and  $Y$  by

$$\text{Cor}(X : Y)_\rho = \max_{\|M\|, \|N\| \leq 1} |\text{Tr}[(M \otimes N)(\rho_{XY} - \rho_X \otimes \rho_Y)]|. \quad (99)$$

Consider some CP and trace-decreasing map  $(\mathbb{I}_X \otimes \mathcal{E}_Y)(\rho_{XY}) = \sum_i E_i \rho_{XY} E_i^\dagger$ . By lemma 9 of Ref. [14], it holds that

$$\text{Cor}(X : Y)_\rho \geq \text{Tr}[M_Y \rho_{XY}] \|\rho_X - \sigma_X\|_1, \quad (100)$$

where  $M_Y = \sum_i E_i^\dagger E_i$  and

$$\sigma_X = \frac{1}{\text{Tr}[M_Y \rho_{XY}]} \text{Tr}_Y [M_Y \rho_{XY}] = \text{Tr}_Y \frac{(\mathbb{I}_X \otimes \mathcal{E}_Y)(\rho_{XY})}{\text{Tr}(\mathbb{I}_X \otimes \mathcal{E}_Y)(\rho_{XY})}. \quad (101)$$

It has been shown that any 1D Gibbs states with a short-range Hamiltonian obey exponential decay of  $\text{Cor}(X : Y)_\rho$  [11], i.e., there exist constants  $c \geq 0$  and  $\xi > 0$  such that

$$\text{Cor}(X : Y)_\rho \leq c e^{-d(X:Y)/\xi}. \quad (102)$$

Choosing  $X = AB_l \dots B_{i+1}$ ,  $Y = B_i$  and  $\mathcal{E}_Y = \tilde{E}_i$  proves lemma 8.  $\square$

Lemma 8 enables us to do an iterative calculation as follows. First we have

$$\begin{aligned} & \left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \tilde{\Lambda}_1(\rho_{AB}^{H_{ABC}}) + (1-p)\tilde{\Lambda}_2(\rho_{AB}^{H_{ABC}}) \right. \\ & \quad \left. + (1-p) \left( \tilde{\Lambda}_3 + \left( \cdots \left( \tilde{\Lambda}_l + \tilde{E}_l \right) \text{Tr}_{l-1} \tilde{E}_{l-1} \right) \cdots \right) \text{Tr}_2 \tilde{E}_2(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq e^{-l/\xi}. \end{aligned} \quad (103)$$

Then we can get

$$\begin{aligned} & \left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \tilde{\Lambda}_1(\rho_{AB}^{H_{ABC}}) + (1-p)\tilde{\Lambda}_2(\rho_{AB}^{H_{ABC}}) + (1-p)^2 \tilde{\Lambda}_3(\rho_{AB}^{H_{ABC}}) \right. \\ & \quad \left. + (1-p)^2 \left( \tilde{\Lambda}_4 \cdots \text{Tr}_4 \tilde{E}_4 \right) \text{Tr}_3 \tilde{E}_3(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq 2e^{-l/\xi}. \end{aligned} \quad (104)$$

We can proceed as above, where at each  $i$ th step, we replace  $\text{Tr}_i \mathcal{E}_i(\rho_{AB}^{H_{ABC}})$  by  $\text{Tr}_i(\rho_{AB}^{H_{ABC}})$  using the triangle inequality. After iterating  $l-1$  steps, we obtain that

$$\left\| \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) - \sum_{i=1}^l (1-p)^{i-1} \tilde{\Lambda}_i(\rho_{AB}^{H_{ABC}}) + (1-p)^{l-1} \tilde{E}_l(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq (l-1)e^{-l/\xi}. \quad (105)$$

Since  $\sum_{i=1}^l p(1-p)^{i-1} + (1-p)^l = 1$ , it follows that

$$\rho^{H_{ABC}} = \sum_{i=1}^l p(1-p)^{i-1} \rho^{H_{ABC}} + (1-p)^l \rho^{H_{ABC}} \quad (106)$$

and thus

$$\left\| \rho^{H_{ABC}} - \sum_{i=1}^l (1-p)^{i-1} \tilde{\Lambda}_i(\rho_{AB}^{H_{ABC}}) + (1-p)^{l-1} \tilde{E}_l(\rho_{AB_l}^{H_{ABC}}) \right\|_1 \quad (107)$$

$$\leq \sum_{i=1}^l p(1-p)^{i-1} \left\| \rho^{H_{ABC}} - \frac{1}{p} \tilde{\Lambda}_i(\rho_{AB_l \dots B_i}^{H_{ABC}}) \right\|_1 + (1-p)^l \left\| \rho^{H_{ABC}} - \frac{1}{1-p} \tilde{E}_l(\rho_{AB_l}^{H_{ABC}}) \right\|_1 \quad (108)$$

$$\leq \{1 - (1-p)^l\} C_2(\beta) e^{-q_1(\beta)l} + 2(1-p)^l. \quad (109)$$

Therefore, by combining Eq. (105) and Eq. (109), finally we obtain that

$$\|\rho^{H_{ABC}} - \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}})\|_1 \leq \{1 - (1-p)^l\} C_2(\beta) e^{-q_1(\beta)l} + 2e^{-|\ln(1-p)|l} + (l-1)e^{-l/\xi} \quad (110)$$

$$\leq C_2(\beta) e^{-q_1(\beta)l} + 2e^{-|\ln(1-p)|l} + le^{-l/\xi}. \quad (111)$$

Here, the probability  $p$  can be bounded as in Eq. (88), and thus we have

$$|\ln(1-p)| \geq \left| \ln \left( 1 - \frac{(1-e^{-1})e^{-2\beta J}}{\left(1 + e^{-\frac{\beta J}{2}} K(\beta)\right)^4} \right) \right| \geq \frac{(1-e^{-1})e^{-2\beta J}}{\left(1 + e^{-\frac{\beta J}{2}} K(\beta)\right)^4} = e^{-\Theta(\beta)}, \quad (112)$$

where the last inequality follows from  $\log(1-x) \leq -x$  for  $x \in [0, 1]$ . If  $\xi = e^{\Theta(\beta)}$ , Eq. (111) can be bounded by

$$2C_2(\beta)le^{-q'(\beta)l} = e^{-q'(\beta)l + \ln(2C_2(\beta)l)}, \quad (113)$$

where  $q'(\beta) \geq e^{-\Theta(\beta)}$ . By exchanging  $l$  by  $d(A, C)$ , we complete the proof.

Note that the all arguments are symmetric under exchanging  $A$  and  $C$ . Therefore, we can obtain a recovery map from  $B$  to  $AB$  with a good accuracy as well.

### C. Proof of Corollary 2

*Proof.* Divide  $C$  into  $C = C_1 \cup C_2 \cup \dots \cup C_m$ , where  $m$  is the maximum number such that  $|C_i| = |B|$  for  $1 \leq i < m$  and each  $C_i$  separates  $C_{i-1}$  from  $C_{i+1}$  (here,  $C_0 \equiv B$ ).

Theorem 1 and Alicki-Fannes inequality imply that

$$I(A : C_i | BC_1 \dots C_{i-1})_{\rho^{H_{ABC}}} \leq |B| e^{-q(\beta) \sqrt{i|B|}} \quad (114)$$

for any  $i \in [1, m]$ . By the chain rule

$$I(A : C | B) = I(A : C_1 | B) + I(A : C_2 | BC_1) + \dots + I(A : C_m | BC_1 \dots C_{m-1}), \quad (115)$$

we have

$$\begin{aligned}
I(A : C|B)_{\rho_{H_{ABC}}} &\leq \sum_{i=1}^m |B| e^{-q(\beta)\sqrt{i|B|}} \\
&\leq \left( |B| e^{-q(\beta)\sqrt{|B|}} + |B| \sum_{i=1}^{m-1} e^{-q(\beta)\sqrt{|B|(i+1)}} \right) \\
&\leq \left( |B| e^{-c\sqrt{|B|}} + \int_1^\infty e^{-q(\beta)\sqrt{|B|x}} dx \right) \\
&= \left( 1 + \frac{2(1 + q(\beta)\sqrt{|B|})}{q(\beta)^2|B|} \right) |B| e^{-q(\beta)\sqrt{|B|}}. \tag{116}
\end{aligned}$$

□

#### D. Proof of Corollary 4

*Proof.* Consider a splitting of the lattice into consecutive regions  $A_1 B_1 C_1 A_2 B_2 C_2 \dots A_k B_k C_k$ , where  $|A_j| = |B_j| = l$  and  $|C_j| = 5\xi l$ , for all  $j$ , with  $\xi \leq \exp(c\beta)$  the correlation length of the Gibbs state. For every  $i$ , data processing inequality and Corollary 2 give

$$I((A_1 B_1 C_1) \dots (A_{j-1} B_{j-1} C_{j-1}) A_j B_{j+1} : C_j | B_j A_{j+1}) \leq C l e^{-c\sqrt{l}}, \tag{117}$$

for constants  $C = \exp(\mathcal{O}(\beta))$  and  $c = \exp(-\mathcal{O}(\beta))$ . By Eq. (7), there is a channel  $\Delta_j : \mathcal{H}_{B_j A_{j+1}} \rightarrow \mathcal{H}_{B_j A_{j+1} C_j}$  s.t.

$$\begin{aligned}
&\|\Delta_j(\rho_{(A_1 B_1 C_1) \dots (A_{j-1} B_{j-1} C_{j-1}) A_j B_j A_{j+1} B_{j+1}}) - (\rho_{(A_1 B_1 C_1) \dots (A_{j-1} B_{j-1} C_{j-1}) (A_j B_j C_j) A_{j+1} B_{j+1}})\|_1 \\
&\leq 2(Cl)^{1/2} e^{-(c/2)\sqrt{l}}. \tag{118}
\end{aligned}$$

Consider the marginal of the Gibbs state on the  $C$  region  $\rho_{C_1 \dots C_k}$ . Using Lemma 20 of [14],

$$\begin{aligned}
&\|\rho_{A_1 B_1 A_2 B_2 \dots A_j B_j} - \rho_{A_1 B_1 \dots A_{j-1} B_{j-1}} \otimes \dots \otimes \rho_{A_k B_k}\|_1 \\
&\leq \sum_{i=1}^k \|\rho_{A_1 B_1 \dots A_i B_i} - \rho_{A_1 B_1 \dots A_{i-1} B_{i-1}} \otimes \rho_{A_i B_i}\|_1 \\
&\leq 2^{|A_i|+2|B_i|} \sum_{i=1}^k \text{cor}(A_1 B_1 \dots A_{i-1} B_{i-1}, A_i B_i)_\rho \\
&\leq k e^{-l}. \tag{119}
\end{aligned}$$

Note that each  $\Delta_i$  act on different sites, i.e. they do not overlap. Then

$$\begin{aligned}
&\|\Delta_1 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1 \tag{120} \\
&\leq \|\Delta_1 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}) - \Delta_2 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k})\|_1 \\
&+ \|\Delta_2 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1 \\
&\leq \|\Delta_1(\rho_{A_1 B_1} \otimes \rho_{A_2 B_2}) - \rho_{A_1 B_1 C_1 A_2 B_2}\|_1 \\
&+ \|\Delta_2 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1 \\
&\leq \|\Delta_1(\rho_{A_1 B_1 A_2 B_2}) - \rho_{A_1 B_1 C_1 A_2 B_2}\|_1 + e^{-l} \\
&+ \|\Delta_2 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1 \\
&\leq 3(Cl)^{1/2} e^{-(c/2)\sqrt{l}} + \|\Delta_2 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1 C_1 A_2 B_2} \otimes \rho_{A_3 B_3} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1.
\end{aligned}$$



The first inequality follows from the triangle inequality, the second from the monotonicity of trace norm under quantum operations, the third from Eq. (119), and the forth from Eq. (118).

Iterating the argument above we find

$$\|\Delta_1 \otimes \dots \otimes \Delta_k(\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}) - \rho_{A_1 B_1 C_1 \dots A_k B_k C_k}\|_1 \leq 3k(Cl)^{1/2} e^{-(c/2)\sqrt{l}} \quad (121)$$

Since  $k \leq n$ , choosing  $l = \mathcal{O}(\log^2(n/\varepsilon))$  gives an error bounded by  $\varepsilon$ .

The depth two circuit is the following: First the state  $\rho_{A_1 B_1} \otimes \dots \otimes \rho_{A_k B_k}$  is constructed in parallel. Then the channels  $\Delta_1 \otimes \dots \otimes \Delta_k$  are applied in parallel.  $\square$

### E. Extension to more general graphs

Our proof for 1D spin chains can be generalized to not only 1D systems but also general graphs with appropriate partitions. For instance, let us consider a tree graph  $G = (E, V)$  with a partition  $ABC$  as depicted in Fig. 3. Since  $G$  is a tree, there is a unique path connecting  $A$  and  $C$ . Then, all spins in  $B$  are classified as (i) spins belonging to the path (ii) descendants of spins on the path (iii) the rest spins which are separated from the path. We can obtain a coarse grained 1D chain by regarding each spin on the path and its descendants as one system, and removing all spins in (iii). Therefore, we can apply the proof in the previous section to this situation as well. Note that the norm of an interaction term connecting spins on the path is irrelevant to the size of the coarse grained spins.

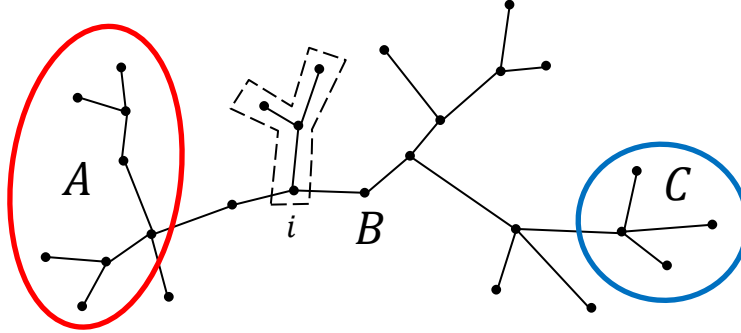


FIG. 3: An example of the region with a partition  $ABC$  for a tree graph. Here,  $B$  is the set of all spins in outside of the circles. In the coarse-grain procedure, the descendants of  $i$  (the spins in the dashed region) can be regarded as one large system.

An important point of the above argument is that the success probability of the recovery map in lemma 7 is bounded by a constant of  $d(A, C)$ . In general cases, we can consider a partition  $ABC$  which cannot be reduced to 1D systems, such as depicted in Fig. 1. Remember that the success probability  $p$  is in the order  $\Omega(e^{\beta\|H_{BM}\|})$ . In the case of Fig. 1,  $H_{BM}$  is the sum of all interactions along the perimeter of  $B^L$ . When considering the repeat-until-success method, the success probability decays too rapidly, and therefore our strategy does not work.

## IV. GIBBS STATE REPRESENTATIONS OF STATES WITH THE APPROXIMATE MARKOV PROPERTY

In this section, we prove Theorem 3, which can be regarded as the converse to Theorem 1, i.e. it states that any quantum approximate Markov chain can be approximated by a 1D Gibbs state

with a short-range Hamiltonian. In the proof, the relationship between the maximum entropy principle and Gibbs form plays an important role. We first review this relationship and then turn to the proof of Theorem 3.

### A. The maximum entropy state and Gibbs states

The maximum entropy principle, introduced by E. T. Jaynes [15, 16], is a useful method to choose an inference under partial information. According to the maximum entropy principle, the most “unbiased” inference is given by the probability distribution with maximum entropy among all distributions satisfying given linear constraints. The solution of this optimization problem is the Gibbs distribution, as can be shown by the method of Lagrange multipliers.

This framework has been generalized to quantum systems (see, e.g., Ref.s [17, 18]). Especially we are interested in the case where the linear constraints are given by reduced density matrices. Let  $\rho_{1\dots n}$  be a quantum state in  $\mathcal{D}(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n)$ . Consider sets of subsystems  $X_1, \dots, X_m$ , with  $X_i \subset \{1, \dots, n\}$ . We define the set  $R_\rho(\mathbf{X})$  by

$$R_\rho(\mathbf{X}) := \{\sigma \in \mathcal{D}(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n) | \sigma_{X_i} = \rho_{X_i}, (1 \leq i \leq m)\}. \quad (122)$$

Therefore  $R_\rho(\mathbf{X})$  is the set of all states with the same reduced states on  $X_i$  as  $\rho$ , for all  $X_i$ . Since  $R_\rho(\mathbf{X})$  is a closed convex set, there exists a unique state such that

$$\sigma_{\max} := \arg \max_{\sigma \in R_\rho(\mathbf{X})} S(\sigma). \quad (123)$$

We call  $\sigma_{\max}$  the maximum entropy state in  $R_\rho(\mathbf{X})$ . Similar to the classical setting,  $\sigma_{\max}$  is given by a Gibbs state of a Hamiltonian with a specific structure. Let us consider the set of Gibbs states  $E(\mathbf{X})$  defined as

$$E(\mathbf{X}) := \left\{ \frac{e^{-H}}{\text{Tr}(e^{-H})} \in \mathcal{D}(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n) \mid H = \sum_{i=1}^m H_{X_i} \right\}, \quad (124)$$

where  $\text{supp}(H_{X_i}) = X_i$ .

Ref. [18] shows that  $\sigma_{\max}$  is the unique element of the intersection of  $R_\rho(\mathbf{X})$  and  $\overline{E(\mathbf{X})}$ , the closure of  $E(\mathbf{X})$  in the reverse-information topology. Moreover, for any  $\omega \in E(\mathbf{X})$ , it is proven that the Pythagorean theorem:

$$S(\rho || \omega) = S(\rho || \sigma_{\max}) + S(\sigma_{\max} || \omega). \quad (125)$$

Since the completely mixed state is always contained in  $E(\mathbf{X})$ , we have

$$\inf_{\sigma \in E(\mathbf{X})} S(\rho || \sigma) = S(\sigma_{\max}) - S(\rho). \quad (126)$$

We will use these formulas in the proof of Theorem 3.

### B. Proof of Theorem 3

We now prove:

**Theorem 3** Let  $\rho_{A_1, \dots, A_n}$  be an quantum  $\varepsilon$ -approximate Markov chain. Then there exists a local Hamiltonian  $H = \sum_i h_{i,i+1}$ , with  $h_{i,i+1}$  only acting on  $A_i A_{i+1}$ , such that

$$S\left(\rho \left\| \frac{e^{-H}}{\text{Tr} e^{-H}}\right.\right) \leq \varepsilon n. \quad (127)$$

*Proof.* Let  $\sigma_{A_1 \dots A_n}$  be the maximum entropy state such that

$$\sigma_{A_i A_{i+1}} = \rho_{A_i A_{i+1}} \quad (128)$$

for all  $i \in [1, n-1]$ . We will show that

$$S(\sigma) \leq S(\rho) + \varepsilon n. \quad (129)$$

The result then follows from Eq. (126).

By strong subadditivity we find

$$\begin{aligned} S(A_1 \dots A_n)_\sigma &\leq S(A_1 A_2)_\sigma - S(A_2)_\sigma + S(A_2 \dots A_n)_\sigma \\ &\leq S(A_1 A_2)_\sigma - S(A_2)_\sigma + S(A_2 A_3)_\sigma - S(A_3)_\sigma + S(A_3 \dots A_n)_\sigma \\ &\vdots \\ &\leq \sum_{i=1}^{n-1} S(A_i A_{i+1})_\sigma - S(A_{i+1})_\sigma \\ &= \sum_{i=1}^{n-1} S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho. \end{aligned} \quad (130)$$

The last equality follows from Eq. (128). By assumption, for every  $i \in [n-1]$ ,

$$I(A_i : A_{i+2} \dots A_n | A_{i+1}) \leq \varepsilon, \quad (131)$$

which can be written as

$$S(A_i \dots A_n) \geq S(A_i A_{i+1}) - S(A_{i+1}) + S(A_{i+1} \dots A_n) - \varepsilon, \quad (132)$$

i.e., strong subadditivity is saturated up to error  $\varepsilon$ . Therefore we have

$$\begin{aligned} \sum_{i=1}^{n-1} S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho &\leq \sum_{i=1}^{n-2} S(A_i A_{i+1})_\rho - S(A_{i+1})_\rho + S(A_{n-2} A_{n-1} A_n) + \varepsilon \\ &\vdots \\ &\leq S(A_1 A_2)_\rho - S(A_2)_\rho + S(A_2 \dots A_n)_\rho + (n-2)\varepsilon \\ &\leq S(A_1 \dots A_n)_\rho - (n-1)\varepsilon. \end{aligned} \quad (133)$$

Combining Eqs. (130) and (133) we get

$$S(\rho \| \sigma) = S(\sigma) - S(\rho) \leq \varepsilon(n-1). \quad (134)$$

By the the discussion in the previous section, there exist a Gibbs state

$$\omega = \frac{1}{Z} \exp \left( - \sum_i H_{A_i A_{i+1}} \right) \quad (135)$$

which satisfies

$$S(\sigma \parallel \omega) \leq \varepsilon. \quad (136)$$

Using the Pythagorean theorem, we completes the proof.  $\square$

## V. ACKNOWLEDGMENTS

Part of this work was done when both of us were working in the QuArC group of Microsoft Research. KK thanks Advanced Leading Graduate Course for Photon Science (ALPS) and JSPS KAKENHI Grant Number JP16J05374 for financial support. We thank Matt Hastings and Michael Kastoryano for useful discussions.

### Appendix A: Another approach to prove Theorem 1

In Sec. III, we employ a perturbative method to obtain a local operator  $O$  determined by the Hamiltonian and a local operator  $V$  such that

$$e^{-\beta(H+V)} \approx O e^{-\beta H} O^\dagger. \quad (A1)$$

The existence of such operator plays a central role in the proof of Theorem 1. In this appendix, we introduce another approach to obtain similar operators, which is based on the previous work by Araki [11]. The main difference between these approaches is the origins of locality of the operator  $O$ . In the perturbative approach, the locality is originated in Lieb-Robinson bounds, which restrict real time evolutions of operators. Instead, in Araki's approach, locality of  $O$  is originated in a restriction on imaginary-time evolutions.

Let us consider a 1D spin chain  $\Lambda = [-n, n]$ , a short-range Hamiltonian  $H$  on  $\Lambda$ , and a local operator  $V$ . We denote the maximum strength of  $H$  by  $J$ . A simple algebra show the following relation holds.

$$e^{-\beta(H+V)} = e^{-\frac{\beta}{2}(H+V)} e^{\frac{\beta}{2}H} e^{-\beta H} e^{\frac{\beta}{2}H} e^{-\frac{\beta}{2}(H+V)} \quad (A2)$$

$$\equiv E_r(V; H) e^{-\beta H} E_r(V; H)^\dagger, \quad (A3)$$

where we denote  $e^{-\frac{\beta}{2}(H+V)} e^{\frac{\beta}{2}H}$  by  $E_r(V; H)$ . By denoting  $V(\beta) = e^{-\beta H} V e^{\beta H}$ ,  $E_r(V; H)$  has another form written as [11]

$$E_r(V; H) = \sum_{n=0}^{\infty} (-1)^n \int_0^{\frac{\beta}{2}} d\beta_1 \int_0^{\beta_1} d\beta_2 \cdots \int_0^{\beta_{n-1}} d\beta_n V(\beta_n) \cdots V(\beta_1). \quad (A4)$$

Actually,  $E_r(V; H)$  can be approximated by a local operator.

**Lemma 9.** [11] *The following statements hold for any region  $X \subset [-n, n]$  and any bounded operator  $V$  with  $\text{supp}(V) = [a, b] \subset [-n, n]$ .*

(i) There exists a constant  $C \geq 0$  depending on  $\beta, J$  and  $\|V\|$  such that

$$\|E_r(V; H_X)\| \leq C \quad (\text{A5})$$

(ii) There exists constants  $C, q \geq 0$  depending on  $\beta, J$  and  $\|V\|$  such that

$$\|E_r(V; H_X) - E_r(V; \lambda H_{X \cap [a-l, b+l]})\| \leq C \frac{q^{1+\lfloor \frac{l}{2} \rfloor}}{(1 + \lfloor \frac{l}{2} \rfloor)!} . \quad (\text{A6})$$

Since  $\log x! \approx x \log x - x$ , the denominator grows faster than the numerator, and thus, the accuracy of the above approximation is exponentially good with respect to  $l$ . Note that similar properties hold for the inverse of  $E_r(V; H_X)$ ,  $E_l(V; H_X) \equiv e^{-\frac{\beta}{2} H_X} e^{\frac{\beta}{2} (H_X + V)}$ . Therefore, by choosing  $V = H_{B^M}$ ,  $E_r(V; H)$  and its local approximation play the same role as  $O_{ABC}$  and  $O_B$  in Sec. III B.

## Appendix B: Comments on approximate Markov property of Gibbs states in higher dimensions

In this paper, we have shown that 1D Gibbs states with short-range Hamiltonians obey approximate Markov property. As we discussed in Sec. III E, our proof cannot be extended for general Markov partition of higher dimensional lattice systems in a straight forward way. In this section, we introduce a conjecture which implies approximate Markov property for any dimensional lattice.

**Conjecture 2.** Let  $H_{ABC}$  be a local Hamiltonian defined on a  $D$ -dimensional lattice with a Markov partition  $A - B - C$ . Then, there exists a bounded hermitian operator  $W_{B'}(\beta)$ , which is supported on  $B' \subset B$ , such that for any  $\beta \leq \beta_c = \mathcal{O}(1)$ , it holds that

$$\left\| \text{Tr}_C \left( e^{H_{ABC}} \right) - e^{H_{AB} + W_{B'}(\beta)} \right\| \leq \mathcal{O} \left( e^{-|B'|} \right) . \quad (\text{B1})$$

In Ref. [19], a similar statement plays an important role to justify the quantum belief algorithm approximates the free energy of a Gibbs state efficiently. The numerical result in Ref. [19] shows our conjecture holds at least for 1D Ising chain with critical transverse field. Also, if the temperature is very high as the same order of  $\|H_{ABC}\|$ , the conjecture can be shown by using the Baker-Campbell-Hausdorff formula.

If Conjecture 2 holds, then there exists a CPTP-map  $\Lambda_{B \rightarrow BC}$  such that

$$\left\| \rho^{H_{ABC}} - \Lambda_{B \rightarrow BC}(\rho_{AB}^{H_{ABC}}) \right\|_1 \leq \mathcal{O}(e^{-|B|}) . \quad (\text{B2})$$

Therefore, the conjecture implies the exponential decay of the conditional mutual information of Gibbs states in any dimension.

To see Eq. (B2), let us consider the following identity.

$$\rho^{H_{ABC}} = (\rho^{H_{ABC}})^{\frac{1}{2}} \left( \rho_{AB}^{H_{ABC}} \right)^{-\frac{1}{2}} \rho_{AB}^{H_{ABC}} \left( \rho_{AB}^{H_{ABC}} \right)^{-\frac{1}{2}} (\rho^{H_{ABC}})^{\frac{1}{2}} . \quad (\text{B3})$$

According to the conjecture, we obtain that

$$(\rho^{H_{ABC}})^{\frac{1}{2}} (\rho_{AB}^{H_{ABC}})^{-\frac{1}{2}} = e^{-\frac{\beta}{2} H_{ABC}} e^{\frac{\beta}{2} H_{AB}} e^{-\frac{\beta}{2} H_{AB}} \left( \text{Tr}_C \left( e^{-\frac{\beta}{2} H_{ABC}} \right) \right)^{-\frac{1}{2}} \quad (\text{B4})$$

$$\approx e^{-\frac{\beta}{2} H_{ABC}} e^{\frac{\beta}{2} H_{AB}} e^{-\frac{\beta}{2} H_{AB}} e^{\frac{\beta}{2} (H_{AB} + W_{B'}(\beta))} \quad (\text{B5})$$

$$= Er(H_{B \cap C} + H_C; H_{AB}) E_l(W_{B'}(\beta); H_{AB}), \quad (\text{B6})$$

where  $H_{B \cap C}$  is the sum of the all interaction terms acting on both  $B$  and  $C$ . Eq. (A6) of Araki's lemma implies both  $Er(H_{B \cap C} + H_C; H_{AB})$  and  $E_l(W_{B'}(\beta); H_{AB})$  can be approximately by operators supported on  $BC$  and  $B$  individually. The accuracy of all approximation is exponentially good with respect to  $|B|$  by choosing  $B'$  as  $\mathcal{O}(|B|)$ . By replacing  $(\rho^{H_{ABC}})^{\frac{1}{2}} (\rho_{AB}^{H_{ABC}})^{-\frac{1}{2}}$  by such local operators, we obtain Eq. (B2).

- 
- [1] Patrick Hayden, Richard Jozsa, Denes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in mathematical physics*, 246(2):359–374, 2004.
  - [2] Ben Ibinson, Noah Linden, and Andreas Winter. Robustness of quantum markov chains. *Communications in Mathematical Physics*, 277(2):289–304, 2008.
  - [3] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.
  - [4] J. M. Hammersley and P. Clifford. Markov field on finite graphs and lattices. 1971.
  - [5] M. S. Leifer and D. Poulin. Quantum graphical models and belief propagation. *Ann. Phys.*, 323(8):1899 – 1946, 2008.
  - [6] W. Brown and D. Poulin. Quantum Markov Networks and Commuting Hamiltonians. *eprint arXiv:1206.0755*, 2012.
  - [7] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac. Area laws in quantum systems: Mutual information and correlations. *Phys. Rev. Lett.*, 100:070502, 2008.
  - [8] Matthew B Hastings. Solving gapped hamiltonians locally. *Physical review b*, 73(8):085115, 2006.
  - [9] M. B. Hastings. Quantum belief propagation: An algorithm for thermal quantum systems. *Phys. Rev. B*, 76:201102, 2007.
  - [10] I. H. Kim. Perturbative analysis of topological entanglement entropy from conditional independence. *Phys. Rev. B*, 86:245116, Dec 2012.
  - [11] H. Araki. Gibbs states of a one dimensional quantum lattice. *Comm. in Math. Phys.*, 14(2):120–157, 1969.
  - [12] E. H. Lieb and D. W. Robinson. The finite group velocity of quantum spin systems. *Comm. Math. Phys.*, 28(3):251–257, 1972.
  - [13] H. Araki. Expansion in banach algebras. *Ann. sci. Ecole Norm. S.*, 6(1):67–84, 1973.
  - [14] F. G. S. L. Brandão and M. Horodecki. Exponential decay of correlations implies area law. *Comm. Math. Phys.*, 333(2):761–798, 2015.
  - [15] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.
  - [16] E. T. Jaynes. Information theory and statistical mechanics. ii. *Phys. Rev.*, 108:171–190, 1957.
  - [17] S. Weis. Information topologies on non-commutative state spaces. *J. Conv. Anal.*, 21(2):339–399, 2014.
  - [18] S. Weis. The maxent extension of a quantum gibbs family, convex geometry and geodesics. *AIP Conference Proceedings*, 1641, 2015.
  - [19] E. Bilgin and D. Poulin. Coarse-grained belief propagation for simulation of interacting quantum systems at all temperatures. *Phys. Rev. B*, 81:054106, Feb 2010.